

Webometria dos repositórios institucionais acadêmicos

Milton Shintaku

Graduado (licenciatura plena) em ciências e habilitação em matemática pelo Centro Universitário de Brasília (Uniceub), pós-graduado *lato sensu* em análise de sistemas pela Universidade Católica de Brasília. Atualmente é tecnólogo no Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) e professor da Secretaria de Educação do Distrito Federal (SEDF) – Brasil.

E-mail: shintaku@ibict.br

Jaime Robredo (* 1927 † 2012)

Doutor em ciências pela Facultad de Ciencias Universidad de Madrid. Madrid – Espanha. Pós-doutorado em ciências químicas na Universidade Politécnica de Delft Holanda.

Dulce Maria Baptista

Doutora em ciência da informação pela Universidade de Brasília. Brasília, DF. Professora da Universidade de Brasília, Departamento de Ciência da Informação e Documentação. Brasília, DF – Brasil

E-mail: dmbp@unb.br

Resumo

O presente estudo oferece um referencial teórico, baseado em estudos de Almind e Ingwersen (1997), Björneborn e Ingwersen (2004), Thelwall, Vaughan e Björneborn (2005), Vanti (2007), entre outros, sobre a webometria, para aplicação em repositórios institucionais acadêmicos, verificando a implementação nessas iniciativas, que, por suas características de agregação da produção científica, torna-se uma representação da comunicação científica nas instituições. O estudo, também, encontra apoio em pesquisadores como Thomas (2007), Dhiman (2010) e Aguillo et al. (2010) sobre estudos webométricos em repositórios. Assim, revelando que métricas aplicadas em repositórios podem apresentar informações importantes sobre a instituição mantenedora do repositório, em relação ao ambiente web.

Palavras-chave

Webometria. Repositórios institucionais acadêmicos. Comunicação científica.

Webometrics of the academic institutional repositories

Abstract

The objective of this paper is to provide a theoretical referential on studies by Almind and Ingwersen (1997), Björneborn and Ingwersen (2004), Thelwall, Vaughan and Björneborn (2005), Vanti (2007), among others, on webometrics, for use in academic institutional repositories, verifying its implementation in these initiatives, which, by their characteristics of added scholar production, become a representation of scholar communication in institutions. The study is also based on research works carried out by Thomas (2007), Dhiman (2010) and Aguillo et al (2010) on webometrics in repositories. The display of metrics applied to repositories may bring out important information on the institution maintaining the repository in regards to the web context.

Keyword

Webometric. Institutional repositories. scholar communication.

INTRODUÇÃO

A Web é, indiscutivelmente, uma fonte de informação importante, mas no que se refere à confiabilidade, no entanto, requer ainda cuidados extras. Nesse caso, os portais acadêmicos, revistas científicas *on-line* e repositórios institucionais são fontes confiáveis, amparados por instituições reconhecidas pela comunidade científica. Assim, essas iniciativas tornam-se fontes de informações confiáveis no ambiente tão vasto da web.

Por esse motivo, as unidades de pesquisa das instituições de ensino superior (IES) (programas de pós-graduação, departamentos, faculdades e institutos) podem criar sua própria revista *on-line*, ou seja, na Web, disseminando resultados de pesquisa

tanto da própria IES quanto de outras, de modo mais abrangente.

Cada vez mais, anais de congressos são acessíveis na web, aumentando consideravelmente o acesso aos artigos deles resultantes. Mas, essas informações geralmente ficam descentralizadas, ou seja, sob a responsabilidade de departamentos, faculdades ou institutos que promoveram o evento.

Com o intuito de agregar a produção científica dispersa pelas unidades, as IES implementam repositórios institucionais. Permitem assim o acesso ao formato integral dessa produção de maneira organizada e com possibilidade de uso de ferramenta de busca em base consolidada. Nesse sentido, os repositórios institucionais acadêmicos tornam-se instrumentos poderosos para as instituições.

Repositórios institucionais acadêmicos tornam-se facilitadores do acesso à produção intelectual das IES. Björk (2005), em seu modelo para a comunicação científica, coloca a disponibilização dos documentos científicos em repositórios como parte da estratégia de facilitar seu acesso. Outros modelos de comunicação científica, como os de Hurd (2000) e Costa (2008), incluem os repositórios como parte do fluxo da comunicação científica. Assim, pode-se notar a importância dos repositórios para a visibilidade das instituições.

Ao considerar os artigos científicos publicados em periódicos e anais de congressos, com a revisão pelos pares, como um dos principais documentos contidos nos repositórios, Costa (2008) revela a proporção da adesão ao formato digital, e por consequência, de sua disponibilização na Web. Assim, mostra a importância da transição do formato impresso para o formato digital.

O fato de os artigos estarem disponíveis em texto completo na Web facilita o acesso e, por consequência, seu fator de impacto, como verificado por Lawrence (2001). Os repositórios institucionais acadêmicos de acesso aberto oferecem a possibilidade de acesso

ao texto integral da maior parte dessa produção, facilitando o acesso e incrementando o fator de impacto Web.

Crow (2002) amplia o impacto dos repositórios, não apenas no fator de impacto Web dos documentos, mas em toda a estrutura da comunicação científica. Bibliotecas acadêmicas, pesquisadores, editores e agências fomentadoras de pesquisa, todos, de forma direta ou indireta são afetados pelos repositórios institucionais de acesso aberto, com a proposta de disponibilização de acervo em acesso aberto.

A representatividade dos repositórios, no entanto, necessita de avaliação baseada em indicadores. Nesse caso, a webometria, com seus métodos e indicadores, pode ser um instrumento valioso nessa averiguação. Nesse ponto, necessita-se adequar as premissas webométricas, que são para a web em geral, aos repositórios.

A webometria, como a métrica relacionada à informação mais recente, requer contextualização à diversa tipologia de iniciativas dispostas na Web. Devido às características distintas de cada iniciativa, além da relação diferenciada entre as iniciativas e seus usuários, o ajuste das métricas para cada iniciativa torna-se necessário para que os resultados representem melhor cada contexto.

Portanto, requiere-se entendimento tanto sobre a métrica quanto a aplicação nos repositórios, que possui característica única como ferramenta da comunicação científica. Inicialmente, a webometria se caracterizava pelo uso em portais acadêmicos, por isso, seus indicadores baseiam-se em características gerais da Web.

O presente estudo tem por objetivo apresentar a webometria, os indicadores, os estudos que envolvem a webometria e, finalmente, o uso dos conceitos e indicadores no âmbito dos repositórios. Essa abordagem objetiva apresentar a métrica em sua forma mais geral e o uso específico em um sistema de informação Web.

WEBOMETRIA

O termo webometria, do inglês *webometrics*, foi utilizado por Almind e Ingwersen (1997) para definir a utilização de métodos infométricos na Web. Com aspectos quantitativos, essa métrica, segundo os autores citados, detém-se na análise estatística da comunicação científica, baseada em uso da rede de computadores, conceituada na infometria e outras métricas. Essa definição ainda se restringe à questão da comunicação científica.

De acordo com Björneborne Ingwersen (2004), webometria é um estudo de aspectos quantitativos, baseados em abordagens infométricas e bibliométricas, da construção e uso de recursos, estruturas e tecnologias na Web, expandindo a abrangência da webometria até as questões tecnológicas. Amplia-se, assim, a cobertura dos estudos webométricos, mas ainda dentro do escopo da ciência da informação.

Thelwall, Vaughan e Bjorneborn (2005) sintetizaram a definição para o estudo quantitativo de todos os fenômenos relacionados com a Web, baseados em métodos originários da bibliometria. Essa definição ampla, no entanto, extrapola a questão acadêmica, já que, segundo os autores, os sítios comerciais são maioria no cenário mundial e podem se beneficiar com esses estudos.

Nota-se que a webometria, como a própria Web, está em evolução, adaptando-se a novas ferramentas e facilidades disponíveis na rede. Se no início da Web as páginas hipertextuais eram uma novidade e todas as páginas possuíam conteúdos estáticos (conhecida como Web 1.0), atualmente as páginas são dinâmicas e montadas no momento da requisição. Nesse contexto, estudos quantitativos sobre a web necessitam adequar-se às novas realidades.

Como métrica, baseada em métodos mais rígidos, a webometria requer constantes avaliações em seus conceitos, métodos e técnicas. Principalmente por ter como objeto de estudo um ambiente em permanente mutação. Assim, há sempre estudiosos

que avaliam e testam as métricas sob o viés da webométrica.

Almind e Ingwersen (1997) vislumbram métodos webométricos para serem aplicados na Web, em substituição aos métodos infométricos, a fim de obter resultados relacionados às características de organização, gerenciamento, uso, produção, disseminação da informação, entre outros. Então, a webometria torna-se uma ferramenta poderosa de apoio a todas as disciplinas que possuem estudos relacionados à Web.

Os estudos da webometria compreendem quatro focos de pesquisa, segundo Thelwald *et al.* (2008), relacionados ao conteúdo, estrutura, uso e tecnologia. Cabe notar que, com essa classificação, cobrem-se todas as possíveis características atuais das iniciativas da Web, independentemente da finalidade. Portanto, o escopo da webometria são os aspectos quantitativos de tudo que se refere à Web.

Nesse ponto, conteúdo difere principalmente no tipo do sítio, visto que informações e serviços são específicos para cada tipologia. Logo, os estudos sobre conteúdo são pertinentes, mas precisam de ajustes dependendo do sítio.

A finalidade do sítio, por sua vez, permite uma classificação, estratificando-o. Eles podem ser classificados como páginas pessoais, governamentais, educacionais, comerciais e outras. Essa divisão, inicialmente pensada na extensão do domínio, como .edu para educação ou .com para sítios comerciais, nem sempre é seguida. Assim a análise deve ser feita criteriosamente pelo conteúdo e facilidades oferecidas e não pela extensão do domínio apresentada na URL.

O tamanho do sítio, por exemplo, é um aspecto contestável. Pode ser verificado por duas características. Tamanho médio das páginas e a quantidade de páginas de um sítio. O tamanho é calculado em unidades de *bits* (ou seu múltiplo *Kbytes*). Páginas multimídias, no entanto, possuem

tamanho bem superior às páginas puramente textuais. Quantidade de páginas de um sítio, no entanto, pode ter indícios melhores, mas necessita de análises mais aprofundadas para ter representatividade.

Dentre os aspectos estudados pela webometria podem-se salientar os estudos relacionados aos *links*. Björneborne e Ingwersen (2001) relatam que a análise dos links, na webometria, é similar aos estudos das citações. Entretanto, pela diversidade e dinamismo da Web, modelos bibliométricos devem ser adaptados à webometria, para possibilitar um estudo mais eficaz, como se cada página disponível na Web fosse um documento, e seus *hiperlinks*, citações.

Os *links* são importantes nos estudos webométricos, pois indicam, principalmente, os relacionamentos entre as páginas. Também revelam como os usuários navegam na Web, o comportamento relacionado à procura de informação (BJÖRNEBORN; INGWERSEN, 2001). Por esses motivos, os *links* se tornam importantes, além de serem de fácil contabilização.

Os estudos baseados nos *links* têm como principal método de coleta os motores de busca Web, que muitas vezes podem apresentar instabilidades, conforme relata Bar-Ilan (2009). Essa metodologia, em muitos casos, requer estudos distintos que permitam verificar sua abrangência, mesmo que a estrutura dos *links* seja padronizada.

O termo *link* substitui o termo português enlace, que, por sua vez, é uma derivação da palavra laço. O termo primitivo enlace (laço) representa a ligação proporcionada pelos *links*, que enlaçam as páginas, criando um relacionamento. Björnborn e Ingwersen (2004) recomendam uma terminologia básica para o relacionamento baseado em links, apresentado no quadro 1. Essa terminologia facilita os estudos e evita dubiedade.

Os *inlinks* e *outlinks* possuem, até certo ponto, o mesmo significado, alterando apenas a direção da origem e destino do *link*. Enquanto nos *outlinks* a página atual remete à outra página, nos *inlinks* a página atual é o alvo de um enlace. Os *selflinks*, por

QUADRO 1
Relacionamentos entre páginas

Relacionamentos através de <i>links</i> sugeridos por Björnborn e Ingwersen (2004)	
<pre> graph TD B[B] --> A[A] A[A] --> C[C] C[C] --> D[D] D[D] --> C[C] A[A] --> A[A] E[E] </pre>	<p>Cada retângulo representa uma página web e cada seta um <i>link</i>. Assim:</p> <ul style="list-style-type: none"> • Página A recebe um <i>inlink</i> da página B • Página A possui um <i>outlink</i> para C • Página E é isolada • Página A possui um <i>selflink</i> • Páginas C e D possuem <i>links</i> recíprocos, <i>interlinks</i> • A página C é acessível desde a página B por um caminho ou <i>Path</i>. • As páginas A e D <i>co-linking</i> a página C

Fonte: Björneborn e Ingwersen (2004)

sua vez, são análogos às autocitações, enquanto os *interlinks* apresentam reciprocidade entre páginas. Os *co-links* representam uma página recebendo *inlinks* de duas outras.

Os *links*, no entanto, possuem maior representatividade se forem externos, ou seja, ligam páginas que pertencem a domínios distintos. Enlaces internos, ou seja, em um mesmo domínio são comuns e não possuem grande representatividade, apenas facilitam a navegação pelos portais. Os *links* externos, por sua vez, representam relações entre os sítios ou portais.

A indexação das páginas por parte dos motores de busca, como os representados pelos *inlinks* e *outlinks*, possui forte influência de algoritmos como os apresentados por Brin e Page (1998). Assim, revela-se que os estudos sobre *links*, no âmbito da webometria, afetam a recuperação em *sites* de busca como o google©, por exemplo.

Os estudos dos *links* permitem contextualizar páginas, sítios, conjunto de páginas ou sítios em relação ao ambiente web. Para tanto, é necessário uma sistematização para esses estudos, e o estabelecimento de indicadores permite que isso seja feito.

QUADRO 2

Quadro resumo dos tipos de links

Quadro resumo dos tipos de links		
Tipo de Link	Direção em relação à página chamada	Explicação
Inlink	Externo → Interno	Internos: São relações internas entre páginas do mesmo portal, ou seja, todos os <i>links</i> que remetem a uma página específica, vista em relação à página chamada.
		Externos: São relações externas entre páginas de portais distintos, ou seja, todos os <i>links</i> que remetem a uma página específica, proveniente de portais distintos e exceto os do mesmo portal, visto em relação à página chamada.
Outlink	Interno → Externo	Internos: São relações internas entre páginas do mesmo portal, ou seja, todos os <i>links</i> , de determinada página que remetem a outras páginas do mesmo portal.
		Externos: São relações externas entre páginas de portais distintos, ou seja, todos os <i>links</i> , de uma determinada página que remetem a páginas de outros portais.
Selflink	Interno → Interno	Páginas: São autochamadas, <i>links</i> que remetem a pontos específicos na mesma página.
		Portais: São iguais aos <i>inlinks</i> internos, <i>links</i> que remetem a páginas de um mesmo portal, servindo para navegação, também conhecidos como <i>links</i> internos.
Interlink	Externo → Interno e Interno → Externo	São enlaces recíprocos, uma página é chamada por outra que remete a essa página, como se fosse um diálogo. Assemelha-se a um <i>loop</i> .
Co-link		<i>Co-inlink</i> : <i>Links</i> provenientes de páginas distintas chamam a mesma página, visto em relação à página chamada. Análogo a cocitação.
		<i>Co-outlink</i> : <i>Links</i> de páginas distintas remetem à mesma página. Análogo a emparelhamento bibliográfico.

Fonte: Björneborn e Ingwersen (2004)

INDICADORES WEBOMÉTRICOS

Vanti (2007) classifica os indicadores webométricos em quatro grupos, e de acordo com suas características, em descritivos, de conectividade, de impacto e densidade. Os estudos podem usar indicadores para descrever sítios ou verificar sua contextualização na Web. O uso de indicadores revela indícios precisos sobre as características e os relacionamentos do sítio.

Os indicadores apresentados a seguir conforme Vanti (2007), baseados em vários estudiosos, tais como Ingwersen (1998), Smith (1999), Thelwald (2000) e Aguillo (2003), entre outros, descrevem os vários aspectos mensuráveis nos recursos da Web.

O indicador descritivo tamanho do *site* indica a quantidade de páginas de um sítio, ou o tamanho em *bytes* dessas páginas. A primeira opção torna-se mais contextualizada com os conceitos atuais, visto que as páginas atualmente são dinâmicas, e assim não possuem tamanho fixo. Nesse caso, a quantidade de páginas pode revelar profundidade, extensão, diversidade e outros aspectos, pois em um sítio as páginas possuem relações internas, seja sobre conteúdo ou pertinência.

Almind e Ingwersen (1997) discutem o tamanho do *site* como um indicador de apoio a outros indicadores. Lawrence e Giles (1999) referem-se à relação com o tamanho total da Web e a capacidade de indexação dos motores de busca. Nesse sentido, enquanto os primeiros autores tratam do indicador tamanho como número de página, o segundo autor do tamanho em *bytes*.

Nota-se que o tamanho em *bytes*, mesmo na comunicação científica, pode ter distorções, já que, em alguns casos, imagens e vídeos, que possuem tamanhos maiores que textos, podem ser anexos de artigos, assim como as bases de dados e *softwares*. Número de páginas, por sua vez, torna-se um indicador mais apropriado aos estudos webométricos.

O indicador de visibilidade se relaciona com os *inlinks* externos. Ele permite verificar a relação de um sítio e o ambiente Web, contabilizando os *inlinks* externos, pois os *inlinks* internos e *selflinks* não representam visibilidade. A coleta desse indicador pode ser efetuada por indexadores como o Google, Yahoo ou Altavista, por exemplo.

A popularidade, por sua vez, tem relação íntima com a visibilidade, mas foca na quantidade de acessos. Esse indicador verifica a quantidade de acessos a um sítio, contabilizando acessos únicos em um determinado período de tempo, ou seja, conta apenas um acesso por endereço IP, mesmo que o mesmo endereço IP visite o sítio por várias vezes. Esse método pode ser controverso, pois um sítio pode ser acessado por várias pessoas em um mesmo IP, como no caso de acessos por *lan houses* ou centros que forneçam internet a comunidades.

O indicador de popularidade, em muitos casos, é disponível em sistemas de estatísticas, como o *googleanalytics* ©, de forma externa ao sistema, ou em ferramentas internas. Torna-se assim o indicador mais acessível aos usuários de sistemas disponíveis na Web.

Fator de impacto Web (Web ImpactFactor – WIF), sugerido por Ingwersen (1998) para ser um indicador semelhante ao fator de impacto das revistas impressas e posteriormente atualizado por Thelwall (2001), é a quantidade de *inlinks* pela quantidade de páginas. Esse indicador possui relevância maior nos estudos webométricos e pode ser aplicado a um sítio, um conjunto de sítios, ou mesmo a sítios de um país. Esse indicador pode revelar quão influente é um sítio no ambiente web.

O indicador de luminosidade relaciona-se com os *outlinks*. Ele revela como se relaciona um sítio e o ambiente web, pois apresenta as conectividades de um sítio a outros. Pode ser útil para verificar comunidade de sítios, pois geralmente os sítios

possuem *outlinks* para outros que têm relação estreita, sejam comerciais, acadêmicas, ou mesmo de assunto.

A densidade média indica a quantidade de *links* que determinado *site* possui, contabilizando todos os *links*, de todos os tipos. Esse montante então deve ser dividido pela quantidade de páginas, e essa relação indica a quantidade média de *links* do sítio. Por esse motivo, traduz como um *site* se relaciona com o ambiente e internamente.

Por fim, a densidade de rede, que possui a finalidade de verificar as conexões nas redes sociais, permite ver os enlaces de mesmo tipo que ocorrem em determinada população, e verificar as relações entre pessoas no ambiente web.

ESTUDOS WEBOMÉTRICOS

Atualmente, a webometria cobre diversos tópicos. Estudos voltados à academia, sítios comerciais ou redes de relacionamento que permitem revelar informações úteis a diversas áreas. É uma ferramenta poderosa e contextualizada com o momento atual, em que a Web se torna parte de cotidiano de muitas pessoas.

Rousseau (1997), em estudo sobre *inlinks*, fez relações entre os domínios mais acessados e a distribuição de Lotka, com estimativa de *selflink* de 30%. Por ainda não possuir um termo mais apropriado, chamou os estudos de infometria ou bibliometria ou cientometria. Entretanto, atualmente, o estudo está no âmbito da webometria.

Vereeland (2000) levantou questões sobre a visibilidade e luminosidades dos sítios das bibliotecas relacionadas com o ensino de Direito, para uma classificação. Observou que a luminosidade dos sítios estudados seguia o padrão 80/20, como na Lei de Bradford, sem no entanto encontrar relação entre a visibilidade e a luminosidade.

Estudos de Thelwalle *et al.* (2008) mostraram a capacidade da webometria em analisar os grupos de pesquisa em países europeus. Os autores verificaram informações sobre os grupos de pesquisa de vários países europeus, como Inglaterra, Alemanha, República Tcheca, Portugal, entre outros. A coleta de dados permitiu a análise sobre a conectividade desses grupos. Detectou-se quais universidades possuem maior número de grupos de pesquisa, visibilidade do grupo, relações entre os grupos, entre outros itens. A webometria se mostrou eficaz nesse tipo de pesquisa.

Os estudos de Fabá-Peres *et al.* (2005) notificaram as relações entre os indicadores webométricos. Apresentaram a interdependência dos indicadores, principalmente dos que pertencem ao mesmo grupo. Analisando sítios relacionados à região de Extremadura, na Espanha, os autores verificaram uma graduação de relacionamento entre indicadores ligados a sítios temáticos.

Pesquisas estatísticas com os *inlinks* devem ser analisadas com cautela, pois necessitam de análise mais profunda. Arakaki e Willett (2009), em estudo sobre a relação de *inlink* e relevância de pesquisa na ciência da informação da Grã-Bretanha, consideram que fatores técnicos e de estrutura dos sites influenciam nos resultados. Muitas das ferramentas existentes não estão apropriadas às pesquisas webométricas. URLs, por sua vez, podem ter diferentes formatos em uma mesma instituição, assim como ter vários níveis, o que dificulta algumas associações.

Assim como os *inlinks* podem indicar fator de impacto web, os *co-links* revelam relacionamento temático. Em pesquisa de *co-link* das unidades de pesquisa da Fundação Oswaldo Cruz, Lang, Gouveia e Leta (2009) revelam que proximidades temáticas aumentam a ocorrência de *co-links*. Essa relação temática também pode ser notada em casos de ocorrência de *interlinks*.

Moscovkin (2010) usou a webometria para verificar o impacto da adoção do movimento do acesso aberto em universidades do consórcio russo-ucraniano-bielorusso. O estudo também apresentou o potencial de ferramentas de busca como o google acadêmico para a geração de índices webométricos. O estudo revelou a possibilidade de avaliação de determinado acontecimento pela webometria.

Fora do âmbito acadêmico, Vaughan e You (2010) fizeram estudo na área comercial de telecomunicações. Baseado na verificação de *co-link* e *co-word*, permitiu ampliar os estudos webométricos para sítios comerciais. Revela-se, assim, que a webometria é aplicável à área comercial, podendo ser usado em *blogs* e outros tipos de páginas, que por serem mais dinâmicas, precisam da delimitação de tempo.

Os estudos webométricos ocorrem em todo o mundo, ainda com certa predominância em estudos em língua inglesa. Por ser um assunto relativamente novo, muitos trabalhos ainda estão por ser divulgados. Sua utilização requer confirmações de uso em várias aplicações e áreas.

WEBOMETRIA NO BRASIL

Os primeiros estudos datam de 2002 (VANTI 2002). Desde então, artigos, teses e dissertações têm sido publicados sobre esse assunto, que pela própria incipiência ainda se apresenta em fase inicial.

Para as teses e dissertações, por exemplo, apenas três teses e duas dissertações estão depositadas na Biblioteca Digital de Teses e Dissertação (BDTD), do Instituto Brasileiro de Informação em Ciências e Tecnologia (Ibict). Esses estudos são apresentados no quadro 3.

Cabe notar que a BDTD não possui 100% das teses e dissertações defendidas no Brasil, mesmo que conte com 97 instituições e mais de 160 mil documentos. Por ter um acervo dinâmico, com adições diárias, a constatação do presente estudo acerca de teses e dissertações na BDTD, sobre webometria, foi feita em junho de 2011.

No âmbito das linhas de pesquisa dos programas de pós-graduação em ciência da informação, apenas duas apresentam, explicitamente, interesse nos estudos webométricos. Estão vinculadas à UFRJ (Comunicação, Organização e Gestão da Informação e do Conhecimento) e à Unesp (Produção e Organização da Informação).

QUADRO 3

Teses e dissertações defendidas no Brasil sobre webometria

Quadro das teses e dissertações sobre webometria				
Autor	Título	Ano	IES	Grau
Vanti, N. A.	Links hipertextuais na comunicação científica: análise webométrica dos sítios acadêmicos latino-americanos em Ciências Sociais	2007	UFRGS	doutor
Gouveia, F. C.	Estudos Webométricos de Associações de Museus e Centros de Ciência	2007	UFRJ	doutor
Lima, J. A. C.	Fluxos de informação no movimento negro brasileiro: do correio Nagô ao virtual	2007	UFF	mestre
Araujo, N. M.	Os planetários e suas relações interinstitucionais no mundo virtual	2010	UFRJ	mestre
Sousa, R. S. C.	Trilhas de comunicação científica : links de postagens de pesquisadores brasileiros nos blogs de ciência	2011	UFRGS	doutor

Fonte: autor

A pequena quantidade de teses e dissertações revela ser este um assunto ainda pouco explorado para pesquisas em nível de pós-graduação. Muito provavelmente pela característica mais técnica, relacionada aos estudos métricos quantitativos, nas ciências sociais e aplicadas, em que está classificada a ciência da informação.

Muitas pesquisas são feitas em universidades e unidades de pesquisa, que não se referem às teses e dissertações. Um referencial de institucionalização de estudos é o Diretório Grupos de Pesquisa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Nesse cadastro estão registrados os grupos de pesquisa certificados pelas instituições, com informações sobre os objetos de estudo e componentes.

No Diretório de Pesquisa do CNPq, apenas três grupos declaram realizar estudos com a webometria. Estão vinculados à Universidade Federal do Rio Grande do Sul (UFRGS) (grupo Comunicação Científica), Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp) (grupo Estudos Métricos em Informação) e Universidade Federal do Rio Grande do Norte (UFRN) (grupo Informação na Sociedade Contemporânea), revelando assim, a diversidade geográfica, mesmo que todas sejam instituições públicas de ensino superior.

Os dados do Diretório de Grupos de Pesquisa mostram que a webometria é um tópico ainda com poucos seguidores. Dos 55 integrantes dos grupos cadastrados no Diretório, 25 são declarados pesquisadores, sendo apenas dois pertencentes a instituições estrangeiras. Dos pesquisadores brasileiros, apenas um não está vinculado à instituição de ensino público.

Faz-se notar, no entanto, a ausência de técnicos nos grupos de pesquisa. Por ser um tópico de cunho técnico e tecnológico, esse dado requer análise mais cuidadosa, pois em muitos grupos vinculados às ciências sociais e humanas é comum a não participação de técnicos.

WEBOMETRIA DOS REPOSITÓRIOS INSTITUCIONAIS ACADÊMICOS

Dhiman (2010) sugere que os repositórios institucionais acadêmicos, como arquivos digitais da produção científica e da produção intelectual de uma instituição, possam fornecer dados adicionais para a pesquisa webométrica relacionada à comunicação científica. Assim, estudos webométricos em relação ao acervo de um repositório fornecerão informações importantes sobre a comunicação científica na era da Internet.

Cabe destacar que os repositórios institucionais acadêmicos são próprios de suas instituições e incluem não somente os artigos de periódicos científicos. Por exemplo, por conter documentos científicos produzidos por várias áreas do conhecimento, os repositórios institucionais acadêmicos reúnem, em seu acervo, documentos oriundos de periódicos científicos, anais de congressos e editoras científicas.

No contexto, a aplicação dos principais indicadores bibliométricos revela não apenas a inserção do repositório no ambiente web, mas a própria inserção da produção intelectual da instituição. Transcende a avaliação webométrica de um sítio para ser a avaliação da produção da instituição, de maneira direta.

Em relação aos indicadores webométricos, no repositório institucional acadêmico, pode ser aplicado para estudar o tamanho do site, visibilidade, popularidade e fator de impacto Web. Podem ser utilizados também para o estudo da luminosidade, densidade média e densidade de rede, mas com menor aplicabilidade em relação aos conceitos webométricos originais.

O tamanho do *site* tem relação direta com o tamanho do acervo do repositório. Note-se que cada objeto digital depositado pode ter uma página, mesmo que dinâmica. Assim o tamanho do *site* é proporcional ao tamanho do acervo. Um indicador que ganha importância não apenas como descrição, mas por

representar o tamanho da produção intelectual disponível no repositório.

A visibilidade tem uma importância maior, como se fosse, realmente, citações. Note-se também que cada recurso disponível no repositório possui uma URL, que pode ser citada. É comum, em citações de documentos digitais, indicar onde o documento pode ser acessado em seu formato completo.

A popularidade, por sua vez, indica o quanto o repositório é acessado, mas um indicativo melhor é a quantidade de vezes que um documento é baixado (*download*). Os acessos vistos pela popularidade representam interesse nos documentos, mas a quantidade de *downloads* possui maior representatividade de possível uso do documento.

O fator de impacto Web, porém, não se mostra um indicativo maior, quanto em outros tipos de sítios. Ele representa a visibilidade mais os relacionamentos internos. É mister citar que os repositórios possuem um quantitativo menor em relação aos *outlinks*, exceto nos casos de repositórios referenciais. Os documentos do acervo, geralmente, não possuem grandes quantidade de *outlinks*.

A luminosidade possui um aspecto diferente das análises tradicionais. Em estudos mais profundos, podem se aproximar de estudos das citações bibliométricas, pois os *outlinks* mais importantes constituem citações de documentos digitais em documentos existentes no acervo. Em repositórios que permitem apenas referências, os *outlinks* podem remeter a outros repositórios, revistas *on-line*, entre outros.

A densidade média pode não ter grande representatividade se vista no geral, mas pode ser decomposta em indicadores mais pontuais nos repositórios. As características do acervo dos repositórios indicariam que a densidade média pode ser vista em função dos tipos de *links*, como a

densidade média de *inlinks*, de *outlinks* e *selflinks*, por exemplo, permitindo revelar o impacto médio dos documentos do repositório (para *inlinks*).

A densidade de rede pode ser útil para indicar as relações entre autores dos documentos pertencentes ao acervo, bem como questões relacionadas à coautoria, apresentando indícios de relações disciplinares e interdisciplinares, ou mesmo, institucionais e interinstitucionais. Quanto às relações disciplinares, podem-se estudar as co-autorias entre autores da mesma área, ou de áreas diferentes. As relações institucionais, por sua vez, podem-se verificar as coautorias institucionais e interinstitucionais.

Outros indicadores, que não estão dentro do escopo da webometria tradicional, podem ser analisados pelas características do repositório. Els indicadores podem revelar outras características não cobertas pelos estudos tradicionais. Tendências ou características de usuários, por exemplo, nem sempre podem ser observadas com precisão pelos indicadores tradicionais.

Os repositórios, geralmente, possuem grandes acervos e os preservam por longos períodos. Em muitos casos, é mais fácil aos usuários utilizar *sites* de busca para acessar o acervo dos repositórios, que, em grande parte, possuem o acervo indexado, que guardar URLs. Se por um lado os sítios de busca fornecem a URL direta do objeto digital desejado, por outro dificultam a contabilização por algumas ferramentas estatísticas automatizadas (indicador de popularidade).

Os padrões dos usuários para acessar o repositório, por exemplo, dão uma informação importante. Há várias formas de acessar um repositório. Os *inlinks* de outros *sites* podem revelar fator de impacto Web, como em citações. Shintaku *et al.* (2009), ao estudar como os usuários acessavam um repositório educacional, verificaram que grande parte acessa o

repositório via indexação do Google, o que pode representar um padrão ou tendência.

Thomas (2007) levanta os principais indicadores de um repositório, dividindo-os em quatro grupos: sobre o desempenho de depósitos, sobre o desempenho das recuperações, sobre o desempenho de uso e sobre o desempenho do impacto. Apresenta, em forma de indicadores, as principais características dos repositórios. Esses indicadores, em muitos casos, são coincidentes com os indicadores webométricos, mas permitem escalonamentos diferentes.

Para os indicadores de desempenho dos depósitos, relaciona informações sobre a participação dos diversos níveis na composição do acervo, desde a participação individual até níveis de programa de pesquisa nacional. Permite, nesse caso, verificar, por estrato, como foi a participação para a composição do acervo do repositório dos pesquisadores, comunidade (departamento, faculdade ou instituto), unidade de informação, institucional e de programa nacional de pesquisa.

Os indicadores de recuperação relacionam-se com as questões de acessos ao acervo do repositório. Possuindo as mesmas estratificações dos indicadores de desempenho diferem, principalmente, no estrato de unidades de pesquisa. Enquanto os indicadores de desempenho centram-se na criação do acervo, os indicadores de recuperação visam à sustentabilidade

e disseminação, tendo um diferencial relacionado à interoperabilidade do repositório.

Quanto aos indicadores de uso, como as mesmas estratificações, centram-se nas questões de satisfação dos usuários finais, em relação às ferramentas disponíveis pelo repositório. Relacionam-se com o acervo e com as facilidades oferecidas pelo repositório.

Por fim, os indicadores de impacto denotam a exposição oferecida pelos repositórios, nos mesmos estratos dos indicadores anteriores. Centrado no impacto dos documentos pertencentes ao acervo do repositório, verifica-se o impacto, destacando, nesse caso, as questões nacionais, que nem sempre são contempladas em estudos, e permitindo verificar o avanço nacional e o montante de conhecimento científico disponível digitalmente na Web.

Nesse contexto, Aguillo et al. (2010) apresentam os indicadores webométricos para uma classificação (quadro 4), para ranquear os repositórios de acesso aberto. Baseado nos motores de busca, gradua os indicadores. Permite-se, assim, criar ranques de repositórios de acesso aberto que podem ser mantidos de forma automática.

Ressalta-se que Aguillo et al. (2010) revelam que, para os repositórios, a visibilidade possui um fator ponderante, mas não único, sendo metade do valor

QUADRO 4

Relação dos indicadores para criação do Rankwebométrico

Quadro de indicadores para rank webométrico		
Indicador	Peso	Descrição
Tamanho	20%	Número de páginas recuperadas por sites de busca baseados em motores de busca, como o Google, Yahoo e outros.
PDF	15%	Quantidade de documentos no formato PDF depositados no repositório. Esse indicador contempla apenas os documentos depositados no repositório. Mesmo que apresente divergências no entendimento. É um subconjunto do indicador tamanho.
Scholar	15%	Quantidade de documentos indexados pelo Google Scholar.
Visibilidade	50%	Quantidade de inlinks, de origem externa ao repositório. Nota-se a importância dada a esse indicador como fator de impacto Web, tornando -se o principal indicador webométrico para ranqueamento.

Fonte: Aguillo et al. (2010).

relacionado à representatividade. Utiliza-se de indicadores webométricos não tradicionais para o ranqueamento, como o PDF e Scholar. As inclusões de indicadores revelam a evolução e a adequação webométrica para os repositórios.

A junção de dois tópicos relativamente recentes, como a webometria e os repositórios institucionais, requer, ainda, estudos mais aprofundados. Iniciativas como o do Open Aire (Europeu) consideram os estudos webométricos para repositórios em fase de desenvolvimento, com a evolução dos dois tópicos.

Em muitos casos, estudos são focados no uso dos repositórios para dar visibilidade à instituição, e aos autores por consequência. Entretanto, estudos baseados na visibilidade do repositório com uso de métricas ainda são incipientes, principalmente no Brasil, que em muitas IES estão em fase de implantação.

CONCLUSÃO

A abrangência do conteúdo do repositório em relação às áreas do conhecimento é um indicador sobre a produção intelectual depositada em um repositório acadêmico. Se para Café *et al.* (2003), um repositório institucional é um conjunto de repositórios temáticos, a abrangência de temas de um acervo tem impacto nos indicadores webométricos, pois permite maiores possibilidades de depósito, recuperação e uso dos documentos.

Thelwall (2001) especulou se as medidas de impacto dos portais acadêmicos podem ter relação sobre a reputação da instituição e seu corpo docente e, conseqüentemente, sobre a qualidade do conteúdo do portal. Nesse sentido, os estudos webométricos nos repositórios permitem verificar esse impacto baseados na produção acadêmica da instituição.

Com base nos indicadores webométricos tradicionais, o quadro a seguir representa um resumo da relação

QUADRO 5

Resumo dos indicadores da webometria tradicionais e os repositórios.

Quadro resumo dos indicadores da webometria tradicionais e os repositórios	
Indicadores	Relação com os repositórios
Tamanho do <i>site</i>	A melhor medida refere-se ao tamanho do acervo e não o tamanho em <i>bytes</i> , visto que cada recurso disponível em um repositório é apresentado em uma página específica.
Visibilidade	São os <i>inlinks</i> recebidos pelo repositório, que podem ser realmente citações aos documentos pertencentes ao acervo, ou mesmo, indicação direta ao documento pela URI (atalhos).
Popularidade	As visitas podem não ter a maior representatividade, mas a quantidade de <i>downloads</i> revela o real interesse pelo documento pertencente ao acervo do repositório.
Fator de impacto Web	Representa a média de links recebidos pelo repositório.
Luminosidade	As medidas mais importantes de luminosidade, em repositórios, referem-se aos <i>outlinks</i> existentes nos documentos contidos no acervo, representando citações a outros documentos digitais.
Densidade média	Pode ser estudada em relação à média dos <i>inlinks</i> e <i>outlinks</i> dos documentos pertencentes ao acervo.
Densidade de Rede	O melhor estudo revela-se quanto às relações de colaboração entre estudiosos de forma interna e externa à instituição, relacionado à coautoria.

Fonte: autor.

desses indicadores e os repositórios institucionais acadêmicos. Esses estudos, centrados nos *links*, visam contextualizar o repositório e a Web.

Com relação aos indicadores não tradicionais à webometria, o quadro traz um resumo desses indicadores e os repositórios institucionais acadêmicos. Eles podem revelar particularidades diferentes daquelas apresentadas pelos indicadores tradicionais, como os diversos níveis com que o repositório se relaciona.

Os indicadores webométricos fornecem um panorama geral da produção intelectual de uma instituição, verificando aspectos que permitem situar a instituição sob vários aspectos, com destaque à visibilidade e impacto. Fornecem indícios que sustentam a viabilidade e importância dessa iniciativa.

Aguillo et al. (2010) agregam ao indicador visibilidade três novos indicadores para compor um conjunto que permite a classificação dos repositórios. Assim, defendem o uso de novos indicadores webométricos, requerendo uma atualização da webometria.

Armbruster e Romary (2010) apontam que a visibilidade é diretamente proporcional à qualidade do acervo. Nesse ponto tem relação com os

indicadores apontados por Aguillo et al. (2010). Entretanto, ressalva-se que os indicadores apenas mostram o resultado de uma estratégia que resulta na visibilidade do repositório. Logo, a visibilidade é o resultado e não o processo.

Nesse contexto, a webometria torna-se uma ferramenta importante para a avaliação e acompanhamento dos repositórios, mesmo que ainda não haja consenso no uso dos indicadores. Por esse motivo, três indicadores tornam-se mais comuns em estatísticas que visam acompanhar a representatividade dos repositórios: a visibilidade, a popularidade e o tamanho do acervo.

No entanto, mesmo com todos os indicadores webométricos, tradicionais ou não, podem ser especialmente úteis para a avaliação dos repositórios, devido às características específicas dessas iniciativas. Nesse ponto, a webometria para os repositórios pode ser enriquecida de outros estudos que resultem em novos indicadores que representem a importância dos repositórios em seu propósito de preservação e como facilitadores do acesso à informação.

Data de submissão: 26-07-2011

Data de aceite: 06-07-2012

QUADRO 6

Resumo dos indicadores da webometria não tradicionais e os repositórios

Quadro resumo dos indicadores da webometria não tradicionais e os repositórios	
Desempenho	Influências na formação do acervo, questões de políticas nos diversos níveis, do pessoal até os programas nacionais de pesquisa.
Recuperação	Relaciona-se à política de acesso aos documentos do acervo, relação às restrições ou ao acesso aberto, refere-se, também, à interoperabilidade do repositório.
Uso	Relaciona-se à satisfação do usuário do repositório quanto ao acervo e facilidades disponíveis.
Impacto	Visibilidade do repositório nas diversas camadas que compõem os usuários do repositório.

Fonte: autor.

REFERÊNCIAS

- AGUILLO, I. F., et al. Indicators for a webometric ranking of open access repositories. *Scientometrics*, v. 82, n. 3, p. 477-486, 2010.
- AGUILLO, Isidro F. Cibermetría: la métrica de la web. In: SEMINÁRIO BUSQUEDA: DEL ARCHIVE A LA RED. Madrid: Residencia de Estudiantes Fundación Francisco Giner de los Ríos, 2003.
- ALMIND, T. C.; INGWERSEN, P. Informetric analyses on the world wide web: methodological approaches to webometrics. *Journal of Documentation*, v. 53, n. 4, p. 404-426, 1997.
- ARAKAKU, M., & WILLET, P. (2009). Webometric analysis of departments of librarianship and information science: a follow-up study. *Journal of information science*, V. 35, n. 2, 143p.
- ARMBRUSTER, C., & ROMARY, L. . Comparing repository types. *International Journal of Digital Library Systems*, v.1, n. 4, p. 61-73, 2010.
- BAR-ILAN, J. Search engine results over time: a case study on search engine stability, *Cybermetrics*, v. 2/3, n. 1, paper 1, 1999. Disponível em: <<http://www.cindoc.csic.es/cybermetrics/articles/v2i1p1.html>>. Acesso em: 10 mai. 2011.
- BJÖRK, B. C. *Scientific communication life-cycle model*, 2005. Disponível em: <<http://oacs.shh.fi/publications/Model35explanation2.pdf>>. Acesso em: Dez. 2010.
- BJÖRNEBORN, L.; INGWERSEN, P. Perspectives of webometrics. *Cientometrics*, v. 50, n. 1, p. 65-82, 2001.
- _____. Toward a basic framework for webometrics. *JASIST*. v. 55, n. 1, p. 1216-1227, 2004.
- BRIN, S.; PAGE, L. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, v. 30, n. 7, p. 107-117, 1998.. Disponível em : <<http://infolab.stanford.edu/~backrub/google.html>>. Acesso em: mai 2011.
- CAFÉ, L. *et al.* Repositórios institucionais: nova estratégia para publicação científica na Rede [Em linha]. In CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 26., 2003, Belo Horizonte, Brasil. Disponível em: <http://dSPACE.ibict.br/dmdocuments/ENDOCOM_CAFE.pdf>. Acesso em: Dez. 2010.
- COSTA, S. M. S. Abordagens, estratégias e ferramentas para o acesso aberto via periódicos e repositórios institucionais em instituições acadêmicas brasileiras. *Liinc em Revista*, v. 4, n. 2, 2008. Disponível em <<http://revista.ibict.br/liinc/index.php/liinc/article/view/281>>. Acesso em: 13 abr. 2009.
- CROW, R. The Case for institutional repositories: A SPARC position paper. Washington, DC: Scholarly Publishing & Academic Resources Coalition, 2002. Disponível em <<http://www.arl.org/sparc/IR/ir.htm>>. Acesso em: fev. 2011.
- DHIMAN, A.K. Webometrics : overview and uses in lis studies for scholarly communication Sixth International. In: CONFERENCE ON WEBOMETRICS, INFORMETRICS AND SCIENTOMETRICS & ELEVENTH COLLNET MEETING, October, 2010, University of Mysore, 2010. p. 19-22.
- FABA-PÉRES, Cet all. Comparative analysis of webometric measurements in thematic environments. *Journal of the American Society for Information Science and Technology*, v. 56, n. 8, p. 779-785, 2005.
- HURD, J. Models of scientific communications system. In: CRAWFORD, S. et al. (Eds), From print to electronic: the transformation of scientific communication. Medford, NJ: Information Today Inc, 1996. p. 9-33
- _____. The Transformation of Scientific Communication: a model for 2020. *Journal of the American Society of Information Science*, v. 51, n. 14, p. 1279-1283, 2000
- INGWERSEN, P. The calculation of web impact factors. *Journal of Documentation*, v. 54, n. 2, p. 236-243, 1998.

- LANG, P. M., GOUVEIA, F. C., LETA, J. Relações intra-institucionais na Internet: um estudo exploratório com base em metodologias webométricas. *Perspect. ciênc. inf.* vol.13, n.3. [online]. 2008. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362008000300009. Acesso em: mai. 2011.
- LAWRENCE, S. *Free online availability substantially increases a paper's impact*. 2001. Disponível em: <<http://www.nature.com/nature/debates/e-access/Articles/lawrence.html>>. Acesso em: maio 2011.
- _____, GILES, C. L. Accessibility and distribution of information on the Web. *Nature*. v. 400, p. 107-109, 1999.
- MOSCOVKIN, V.M. The webometric estimate of the publication activities of universities: the influence of the belgorod declaration. *Scientific and Technical Information Processing*, v. 37, n. 1, p. 49-54, 2010.
- ROUSSEAU, R. . Sitation: an exploratory study. *Cybermetrics*, n. 1 v. 1. 1997. Disponível em: <http://cybermetrics.cindoc.csic.es/articles/v1i1p1.html>. Acesso em: abr. 2011.
- SHINTAKU, M. et all, Banco Internacional de Objetos Educacionais (BIOE): uma iniciativa para disseminação de objetos educacionais pela internet. *Revista Intercâmbio dos Congressos de Humanidades*, 2009. Disponível em: <http://unb.revistaintercambio.net.br/24h/pessoa/temp/anexo/1/421/749.pdf>. Acesso em: abr. 2011.
- SMITH, A. G. (1999). A tale of two web spaces: comparing sites using Web Impact Factors. *Journal of Documentation*, v. 55, n. 5, p. 577-592.
- THELWALL, M . Extracting macroscopic information from web links. *Journal of the American Society for Information Science and Technology*. v. 52, n. 13, p. 1157-1168, 2001.
- _____. Introduction to webometrics: quantitative web research for the social sciences. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, San Rafael, CA: Morgan & Claypool, 2009.
- _____, et al. . Assessing the international web connectivity of research groups, *Aslib Proceedings*, v. 60, n. 1, p.18 – 31, 2008.
- _____, VAUGHAN, L. BJÖRNEBORN, L. Webometrics. *Annual Review of Information Science and Technology*, n. 39, p. 81–135, 2005.
- Thelwall, M. (2000). Web impact factors and search engine coverage. *Journal of Documentation*, v. 56, n. 2, p. 185-189.
- THOMAS, G. M. E. Evaluating the impact of the Institutional Repository, or positioning innovation between a rock and a hard place. *New Review of Information Networking*, v. 13, n. 2, p. 133-146..
- VANTI, Nadia . Links hipertextuais na comunicação científica: análise webométrica dos sítios acadêmicos latino-americanos em Ciências Sociais. Porto Alegre, 2007. 299 f. Tese (Doutorado) – Programa de Pós-Graduação em Informação e Comunicação. Universidade Federal do Rio Grande do Sul, Porto Alegre, 2007.
- _____. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. *Ciência da Informação*, Brasília, v. 31, n. 2, p. 152-162, maio/ago. 2002.
- VAUGHAN, Liwen; YOU, Justin. Word co-occurrences on webpages as a measure of the relatedness of organizations: A new Webometrics concept. *Journal of Informetrics*, v. 4, n. 4, p. 483-491, 2010.
- VEREELAND, R. C.. Law libraries in hyperspace: A citation analysis of world wide web sites. *Law library journal*, v. 92 n. 1, pp. 9-25. 2000. Disponível em: http://69.63.136.221/main-menu/Publications/llj/LLJ-Archives/Vol-92/pub_llj_v92n01/2000-2.pdf. Acesso em: abr. 2011.